

Fusion based estimation of the a-priori probability distribution of unknown non-stationary processes

Marek Junghans

*Institute of Transportation Systems
German Aerospace Center (DLR)
Berlin, Germany
marek.junghans@dlr.de*

Andreas Leich

*Institute of Transportation Systems
German Aerospace Center (DLR)
Berlin, Germany
andreas.leich@dlr.de*

Abstract—Non-stationary processes can be hard to handle, particular if one would like to know their characterizing time dependent probability functions. In this paper the a-priori probability distributions of unknown non-stationary processes are estimated with different combinations of weakly coupled sensors. For quantification of the unknown a-priori probabilities Bayesian Networks (BN) are adopted for data fusion and Dirichlet functions are applied on non-stationary, time-dependent maximum likelihood (ML) parameter learning. In several experiments the adaption of the non-stationary a-priori probability density functions is shown and the accuracy of data fusion regarding the underlying process variables with different characteristics are determined quantitatively. It is shown that the proposed algorithm can improve data fusion in case conditions for specific process and sensor characteristics are met.

Index Terms—Bayesian Networks, Maximum-likelihood parameter learning, non-stationary processes

I. INTRODUCTION

In stationary stochastic processes the statistical properties of the underlying and the process characterizing signal(s) do not change over time and thus, are easy to handle. In particular, if someone wants to know more about the properties of a stationary process (e.g. mean, variance, probability density function) specific learning methods can be applied. In contrast, if a stochastic process turns out to be non-stationary, it can be rather challenging or even impossible to quantify its properties. Thanks to modern methods of machine learning and system theory we can find out a lot about highly non-stationary processes, particularly about their underlying a-priori probability distributions, whose determination will be the focus of this paper. In general, stochastic processes can be modelled as dynamical systems with input variables, process variables (which are usually hidden) and output variables (which are usually measured). In order to extract detailed information about the stochastic process, either the transfer function of the system is needed or the system's characteristics can be provided by suitable physical models that describe the behavior of the dynamical system. For this purpose sensors can be used. Clearly, sensors are imperfect dynamical systems too that are influenced by their by environmental factors (e.g. physical conditions) as well as their inner characteristics (e.g. thermal noise).

Daily traffic is a typical example of stochastic processes: It is the result of the random mobility needs of each individual road user depending on the road users' circadian rhythms and needs as well as motion behaviors like driving, accelerating, braking, stopping, waiting, etc. Although the traffic process can be characterized by stationary and even ergodic proportions (e.g. free flow situations), traffic is non-stationary in many cases (e.g. traffic breakdowns due to car crashes), when its density increases and the speeds go down. Further, traffic follows some recurrent patterns, e.g. morning/afternoon peak hours, traffic jams, that do happen (almost) every day. But the statistical composition of traffic, i.e. the types of road users, changes from minute to minute, crashes between road users happen at random and are almost impossible to predict, as well as affections by calculable and unplanned major events. Additionally, traffic is influenced by traffic light control, which complicates the prediction and extrapolation of the traffic state from a certain place to another at some specific time. All in all traffic is a highly non-stationary process.

There are many different types of sensors available, which are capable of measuring certain aspects of traffic, e.g. camera sensors. Although there are advanced methods applicable to handle camera data under the most complex conditions, e.g. Neuronal Networks (NN) of different designs, for instance (Deep) Convolutional NN (CNN, DNN), Recurrent NN (RNN) [1], [2], etc. it is obviously more difficult for a video sensor to detect, classify and even track objects in case of difficult weather and/or illumination conditions, such as heavy snowing. On the basis of this example it is clear that sensors always provide biased data of the underlying process; and thus applying biased sensory data for estimating the underlying process without taking into account the inner and environmental influences will also increase the bias of the knowledge about the unknown process. In [3], [4] probabilistic sensor models based on Bayesian Networks (BN) are applied to reduce bias of measured data by modelling the sensors' inner an environmental influences. Accordingly, fusing de-biased data of different (or homogeneous) sensors improves quality and expressiveness of the measured data with regard to the unknown underlying process.

At this point the question arises: *To what extent can we take advantage of quantifying our knowledge about the unknown*

underlying non-stationary process if its a-priori probability distribution is modelled time-dependent? In this paper we will try to answer this question. For this purpose an algorithm is developed that allows to estimate the a-priori probability distribution of an unknown non-stationary process taking into account weakly coupled influenced sensors. On the basis of three experiments with different process characteristics it is shown that the estimation of the time-dependent a-priori probability distribution can be improved in case the underlying process is featured by specific characteristics. Additionally, we can show that the combination of different sensors may be problematic regarding the task in question.

The rest of the paper is structured as follows: In section II the problem in question is formulated and the required variables are defined. In section III the concept of Bayesian Network based data fusion of weakly coupled sensors is introduced. Then, in section IV the well-known concept of Dirichlet distributions for parameter learning is applied for non-stationary, time-dependent learning of the a-priori probability distribution. In section V the experimental setup and the obtained results are presented. Finally, in section VI the paper is summarized and future prospects are presented.

II. PROBLEM FORMULATION

The estimation of the a-priori probability distribution of an unknown, non-stationary, stochastic process is a task, which may be arbitrarily complex. In the following paragraphs the necessary variables are introduced and the problem described. In section II-A the unknown process is introduced and in II-B extended by a Dirichlet source for learning purposes. In section II-C the measuring process is defined, which is extended by the consideration of several sensors in the sense of data fusion in II-D. Finally, in section II-E the process knowledge update described.

A. Unknown process X

Let us consider an unknown, non-stationary stochastic process $X(t)$ ¹. X shall be unobservable directly, i.e. its realizations $x \in X$ are hidden, but can be observed indirectly by sensors attached to X . Let us further assume there is a mapping function Υ_X , which transforms the realizations x from time domain to probability domain yielding the probability density function $P(x) = \Upsilon_X(x)$ with discrete realizations $x = \{x_1, \dots, x_{N_x}\}$. Then, $P(x)$ models the a-priori probability distribution of X .

Since X is defined as a non-stationary process, at least the expectation value $\mathbb{E}(X)$ and the variance $\text{Var}(X)$ change over time. However, to facilitate the problem in question we assume X to have quasi-stationary properties within a time-dependent batch of size $C(t) \in \mathbb{N}$. Clearly, if C is known, at least $\mathbb{E}(X|C)$ and $\text{Var}(X|C)$ remain constant and we are allowed to handle X like a stationary process applying classical statistical methods for evaluation and learning. Although the size of C can vary, we can qualitatively state that in case

¹For reasons of convenience we will drop time t in cases, where it is not explicitly needed.

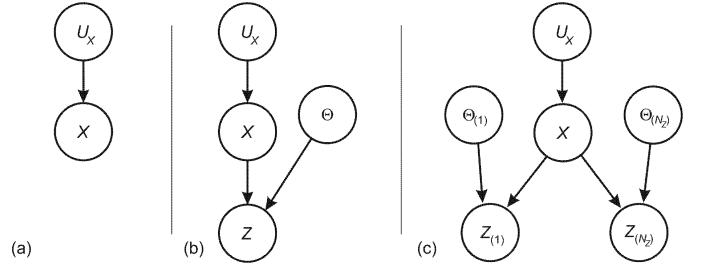


Fig. 1. Modelling different cause-effect relationships in causal graphs: (a) non-stationarity of process X by U_X ; (b) in addition to (a) the affection of the measurement process Z by Θ ; (c) N_Z weakly coupled, affected sensors (for sake of clarity only the two measuring processes $Z_{(1)}$ and $Z_{(N_Z)}$ and their influencing processes $\Theta_{(1)}$ and $\Theta_{(N_Z)}$ are shown)

C is (very) big, X tends to be (more) stationary, whereas if C is (very) small, X tends to be (highly) non-stationary. A convenient way to model non-stationarity of X is to add another stochastic process variable, which shall be defined as $U_X(t)$. U_X characterizes the change of X 's properties over time. X can be written as $X = X(U_X)$ or, equivalently, as a conditional process $X|U_X$ and it can be understood as a continuous switch to either increase stationarity or decrease it. Making it simpler, if $U_X = \emptyset$, X shall be stationary, in all other cases non-stationary. The realizations of U_X are $u_X \in U_X$. Mapping $x|u_X$ from time domain to probability domain yields the new a-priori probability distribution $P(x|u_X)$. The cause-effect relationship between U_X and X is shown in fig. 1(a) as graphical model.

B. Adding Dirichlet source to X

A convenient way to quantify $P(x|u_X)$ is to apply the concept of Dirichlet distributions (see [5]–[7]). A Dirichlet density distribution of a variable H with its realizations $h = \{h_1, \dots, h_{N_x}\}$ is given by:

$$P(h) = \text{dir}(h, c) = \frac{\Gamma(C)}{\prod_{i=1}^{N_x} \Gamma(c_i)} \prod_{i=1}^{N_x} h_i^{c_i-1}. \quad (1)$$

In eq. (1) c characterizes how frequently event $X=x$ occurs, whereas all counts of all events sum up to $C = \sum_i c_i$; $0 \leq c_i \leq 1$; $\Gamma(\cdot)$ is the Gamma density distribution function. Let us assume the unknown process X is influenced by a Dirichlet source H . Then, the affection of X by H leads to an a-priori probability distribution, which shall be defined as $P(x|h, u_X) = h$ with $h > 0$ and $\sum_{\forall h} h = 1$. $P(x|u_X)$ then is:

$$P(x|u_X) = \int P(x|h, u_X) \cdot P(h) dh = \int h \cdot P(h) dh = \mathbb{E}(H) \quad (2)$$

Setting $P(h) = \text{dir}(h, c)$ and computing eq. (2) yields the (stationary) quantification of a countable a-priori distribution density function $P(x|u_X) = c/C$, which reflects the expectation value of process H . Note, that $c = (c_1, \dots, c_{N_x})^T$ is a vector of counts (so called hyper parameter [7]) of each realization x . If we were able to observe X , we could

determine $P(x|u_X)$ directly by counting the events x (see also [6], [8]):

$$P(x|u_X) = \mathbb{E}(H) = \int h \cdot \text{dir}(h, c) dh = \frac{c}{C}. \quad (3)$$

In case we have identified a realization $x = x_i$, $1 < i < N_X$ we can calculate the influence of x_i on the Dirichlet source H by applying Bayes' rule [6]:

$$P(h|x, u_X) \propto P(x|h, u_X) \cdot P(h) = \text{dir}(h, c_i + 1) \quad (4)$$

which leads to the updated expectation value $\mathbb{E}(H|X)$ by adding just one more sample to the counts $c_i := c_i + 1$ due to the fact that x_i occurred:

$$\mathbb{E}(H|X) = \frac{(c_1, \dots, c_i + 1, \dots, c_{N_X})^T}{C + 1}. \quad (5)$$

Eq. (5) shows that an identification of the unknown x_i leads to an integer increment of the associated counter value $c_i := c_i + 1$ and thus to an increment of the overall size of all values contained, i.e. $C := C + 1$, too.

C. Measuring process Z

The realizations of X are hidden and we assume there is a sensor available to measure x . The observation process $Z(t)$ is usually nonlinear and can also be characterized by a non-stationary process. The measurement results are projections of x to z yielding the realizations $z \in Z$. The measuring process can be described by the observation eq. (6):

$$z = \zeta(x, \theta) \quad (6)$$

In eq. (6) the functional relation ζ can be described as a container for all methods and technical hardware the sensor needs to measure x taking into account influencing factors θ , e.g. the physical measuring principle, the number an intensity of environmental influences, conditions of operation, etc. A simple, but frequently used version of eq. (6) is $z = x + \epsilon$, where ϵ is modelled by additive noise, which may follow some error distribution. A more complex version of eq. (6) may contain more influencing factors, multiplicative noise and dynamic errors, which is not shown here, but the reader is referred to [9] (in German). It seems reasonable to model such influences by stochastic influence variables. Therefore we define $\Theta(t)$ as variable for several stochastic processes affecting the observation process Z . Consequently, eq. (6) also contains the sensory influence parameter θ with the realizations $\theta \in \Theta$. The cause-effect relationship is shown in fig. 1(b) as graphical model.

Mapping the measurement results z from time to discrete probability domain by a mapping function Υ_Z yields the conditional probability density function $P(z|x, \theta) = \Upsilon_Z(z(x, \theta))$. The resulting conditional probability distribution $P(z|x, \theta)$ is called sensor likelihood, which probabilistically describes the projection of the realizations x onto the realizations z . Clearly, $P(z|x, \theta)$ quantifies how the sensor statistically behaves with regard to the process variables x and the sensor affections θ .

D. Data fusion

Although many definitions of the term *Data Fusion* and many accurate descriptions exist (see [10]–[16] for instance), they all have in common that data of $N_Z \in \mathbb{N}$ sensors are combined in a way so that the resulting knowledge of the unknown process is *better* than without. The term *better* shall be understood in the sense of a greater accuracy, precision, reliability, completeness, etc. The graphical model in fig. 1(c) shows the cause-effect relationships among all the processes and particularly the conditional independence of the observation processes $Z_{(1)}$ and $Z_{(N_Z)}$ in case of weakly coupled sensors.

Let us assume there is an optimal set of methods and techniques combined in the functional ψ , which are capable of optimally fusing the measurement results of N_Z sensors. Then we can write the fusion equation with $\mathbf{z} = \{z_{(1)}, \dots, z_{(N_Z)}\}$ and $\theta = \{\theta_{(1)}, \dots, \theta_{(N_Z)}\}$ as:

$$x \approx \hat{x} = \psi(z_{(1)}, \dots, z_{(N_Z)}; \theta_{(1)}, \dots, \theta_{(N_Z)}) = \psi(\mathbf{z}, \theta) \quad (7)$$

In eq. (7) \hat{x} is the optimal estimation result of x and $z_{(i)}$ is the measurement result of observation process $Z_{(i)}$ taking into account the influence $\theta_{(i)}$.

E. Process knowledge update by data fusion

In order to find an optimal estimate \hat{x} in the single sensor case as described by eq. (6) we have to invert the functional relation ζ . Instead, in case of N_Z sensors we must solve eq. (7). Therefore, we obtain the following two equivalent equations with their functional relations ζ^{-1} and ψ .

$$\hat{x} \approx x = \zeta^{-1}(z, \theta) \quad (8)$$

$$\hat{x} \approx x = \psi(\mathbf{z}, \theta) \quad (9)$$

Coming back to our Dirichlet source H we added in subsection II-B to describe parameter learning by the application of Dirichlet density distributions, we can now compute the influence of each measurement result $z \in \mathbf{z}$ on H and X by solving the Kolmogorov-Chapman equation taking into account the marginalization of x by applying the total probability theorem:

$$\begin{aligned} P(h|\mathbf{z}, u_X, \theta) &= \sum_{\forall x} P(h, x|u_X, \mathbf{z}, \theta) \\ &= \sum_{\forall x} P(h|x, u_X) \cdot P(x|\mathbf{z}, \theta) \end{aligned} \quad (10)$$

Let us now consider the measurement results \mathbf{z} for updating $P(x)$. Following the scheme of updating the Dirichlet source H by the eqs. (4) and (5) as well as applying eq. (10) leads to the desired quantification:

$$P(h|\mathbf{z}, u_X, \theta) = \text{dir}(h, c + P(x|\mathbf{z}, \theta)) \quad (11)$$

which emphasizes that c is not updated by an integer, but by a whole probability density function instead [7]. Computing $\mathbb{E}(H|Z_{(1)}, \dots, Z_{(N_Z)}) = \mathbb{E}(H|\mathbf{Z})$ yields the updated a-priori probability density function taking into account all the measurement results of all measuring processes:

$$P(x|u_X) := \mathbb{E}(H|\mathbf{Z}) = \frac{c + P(x|\mathbf{z}, \theta)}{C + 1} \quad (12)$$

Eq. (12) is only valid in case of a stationary process X , i.e. $U_X = \emptyset$. Further, each measurement result z updates and sharpens the a-priori probability distribution due to incrementing the number of overall samples of C to $C := C+1$, etc. In section IV a parameter learning method is introduced to handle a-priori updates in case of a non-stationary process.

III. BAYESIAN NETWORK BASED DATA FUSION

Bayesian Networks (BN) are commonly used for fusing different homogeneous or heterogeneous sensors to increase the quality of measured data and to increase the detection horizon. As we know from many publications, for instance in [17]–[22], BN are powerful to provide reliable and accurate fusion results by inferring the available measurement results of the applied sensors taking into account their environmental affections.

In the following BN are briefly described (III-A) and applied on weakly coupled sensors (III-B). The resulting fusion equation is provided taking into account the unknown stochastic non-stationary process, the measuring and influence processes.

A. BN briefly described [6], [23]

A BN is a graphical formalism of handling and processing uncertain and incomplete knowledge in causal reasoning. BN consist of a set of discrete random variables (nodes) and a set of directed links. Each node is described by a set of mutually exclusive states. Some of the nodes are connected with other nodes by directed links. These links characterize the conditional dependencies among the connected nodes. The cause-effect relationships in the BN are quantified by conditional probability density functions to each single node. The nodes together with the links form the directed acyclic graph (DAG). A DAG is a BN if the nodes of the DAG satisfy the structural Markov condition, i.e. they are conditionally independent on its non-descendents given its parents. For detailed information on structure, computation and inference of BN the reader is referred to [6] and [23].

B. Weak Data Fusion with BN

In [24] mainly two data fusion concepts are distinguished: *weak* and *strong data fusion*. In case of several *weakly* coupled sensors their outcomes are conditionally independent on the underlying process. In case of *strong* data fusion, one or more measuring processes are dependent on one or more other measurement processes as well. In this paper only the concept of *weak data fusion* is considered.

In case all necessary conditional probability density functions, i.e. a-priori probability distributions and the sensor likelihoods, are quantified, the DAG in fig. 1(c) is a BN for fusing the realizations of the weakly coupled measuring processes $Z_{(i)}$, $1 < i \leq N_Z$, taking into account process X and the sensors' affections $\Theta_{(i)}$, $1 < i \leq N_Z$. For reasons of clarity, only two measuring processes $Z_{(1)}$ and $Z_{(N_Z)}$ and two

influence nodes $\Theta_{(1)}$ and $\Theta_{(N_Z)}$ are depicted. The resulting JPD, which satisfies the Markov condition, is:

$$P(x, u_X, \mathbf{z}, \theta) = P(x|u_X) \cdot \prod_{i=1}^{N_Z} P(z_{(i)}|x, \theta_{(i)}) \quad (13)$$

In eq. (13) the term $P(x|u_X)$ characterizes the a-priori probability density by taking into account the (non-)stationarity of X by the “switch” u_X . The term $P(z_{(i)}|x, \theta_{(i)})$ is the likelihood of the i th measuring process considering the influences $\theta_{(i)} \in \Theta_{(i)}$. The a-posteriori probability density $P(x|u_X, \mathbf{z}, \theta)$ with $\alpha^{-1} = P(\mathbf{z}|\theta)$ then is:

$$P(x|u_X, \mathbf{z}, \theta) = \alpha \cdot P(x|u_X) \cdot \prod_{i=1}^{N_Z} P(z_{(i)}|x, \theta_{(i)}) \quad (14)$$

Eq. (14) can be solved by applying inference methods introduced in detail in [6], [23]. By applying an adequate parameter estimator, e.g. maximum-a-posteriori (MAP) or probability wheel (PW) [5], on eq. (14) the fusion equation (7) or equivalently the eqs. (7) and (9) are solved. In case MAP is used we obtain the optimal estimate $\hat{x} \approx x$:

$$\hat{x} = \arg \max_x P(x|u_X, \mathbf{z}, \theta) \quad (15)$$

IV. LEARNING THE A-PRIORI PROBABILITY DISTRIBUTION

For learning parameters in BN there are powerful algorithms available in case of an unknown stationary process. For instance, in [7] a ML (maximum likelihood) algorithm was developed. In contrast, not much is known in case of estimating unknown, stochastic non-stationary a-priori probability distributions. The problem of the slow convergence of the ML algorithm in [7] was solved by developing a voting EM (expectation maximization) in [25] considering dynamic learning rates. In [26] a combination of ML and Voting EM was proposed. In this paper a simple sequential method for learning a stochastic, unknown and non-stationary a-priori distribution is introduced in section IV-A, which is based on the ML algorithm in [7] and on investigations made in [27] and [28]. In section IV-B some statements on the proposed algorithm are given.

A. Non-stationary case

Let us consider a non-stationary process X , i.e. $U_X \neq \emptyset$, which has quasi-stationary properties within a batch of data samples $C(t)$. Although C may change as time t increases we assume that the size of $C(t) = C$ remains constant. Thus, C measurement results can be taken into consideration. Any new measurement will lead to the eqs. (11) and (12), respectively.

At some time $t > t_0$, when the latest fusion result of the sensors has been obtained, eq. (12) leads to the (temporal) a-priori update, which is the same as in the stationary case, but however, it is not the entire truth of determining the desired a-priori probability function:

$$\begin{aligned} P(x|u_X, t) &= E(H|\mathbf{Z}, t) \\ &= \frac{c(t-1) + P(x|u_X, \mathbf{z}, \theta, t)}{C+1} \end{aligned} \quad (16)$$

The count vector $c(t-1)$ shall indicate the fusion result one time step before the fusion result occurred, and is thus updated to $c(t) = c(t-1) + P(x|u_X, \mathbf{z}, \theta, t)$. Accordingly, the whole batch size increases by 1. Now we have to drop the oldest fusion result of this batch from being taken into account. Let t_0 be the time when the oldest fusion result was obtained and let $P(x|u_X, \mathbf{z}, \theta, t_0)$ be the according a-posteriori probability function. Then we obtain:

$$P(x|u_X, t) = E(H|Z_{(1)}|_{t_0}^t, \dots, Z_{(N_Z)}|_{t_0}^t) = \frac{c(t-1) + \varepsilon(t) - \delta(t)}{C}. \quad (17)$$

For reasons of simplicity we substituted the above mentioned latest and oldest a-posteriori probability density functions by the variables ε and δ respectively:

$$\varepsilon(t) = P(x|u_X, \mathbf{z}, \theta, t) \quad (18)$$

$$\delta(t) = P(x|u_X, \mathbf{z}, \theta, t_0) \quad (19)$$

Note, that $t_0 := t_0(t)$ is time-dependent. When time increases, let's say by one, t_0 (usually) increases by one too. In general, the count vector $c(t)$ can be rewritten as follows taking into account the size C :

$$c(t) = c(t-1) + \varepsilon(t) - \delta(t) \quad (20)$$

The a-priori knowledge can be updated immediately when fusion results are available. This leads to a sequential learning process. To avoid fluctuations it is reasonable to update $P(x|u_X, t)$ only if a certain threshold has been exceeded. The Kullback-Leibler divergence d_{KL} of the a-priori probability density functions with the threshold T_{KL} is an adequate metric:

$$d_{KL}(p, q) = \sum_{\forall x} p(x) \cdot \log \frac{p(x)}{q(x)} \quad (21)$$

Applying eq. (21) on our problem in question leads with $p = P(x|u_X, t_1)$ and $q = P(x|u_X, t_0)$ to:

$$P(x|u_X) := \begin{cases} P(x|u_X, t_1) & d_{KL}(p, q) > T_{KL} \\ P(x|u_X, t_0) & d_{KL}(p, q) \leq T_{KL} \end{cases} \quad (22)$$

B. Statements on the algorithm

Concerning the update of the unknown, stochastic, non-stationary a-priori probability distribution on the basis of N_Z affected sensors $Z_{(1)}, \dots, Z_{(N_Z)}$ it can be stated:

- $P(x|u_X, t)$ can be computed by considering the unknown, non-stationary process X as quasi-stationary, since it provides stationary and thus, computable properties of X by a constant number of measurement results. Therefore, on the basis of a stationarity check of X , a value of C has to be chosen, which quantifies the amount of stationarity within non-stationarity. Note that C must not be too big, which leads to smoothing out essential information; and not too small to avoid stochastic fluctuations. Also, the rarest realization should be taken into account. Consequently, it is a difficult question how to determine a decent size of C .

- X is observed by a set of N_Z weakly coupled sensors, whose measurement results are combined probabilistically in the sense of weakly coupled data fusion. The inference of BN with regard to the problem in question leads to an a-posteriori probability density $P(x|u_X, \mathbf{z}, \theta, t)$ taking into account all the measurement results and influence parameters at each time t .
- As time increases and new fusion results occur, the computed a-posteriori probability density functions serve for updating the event count vector c : First, the newest fusion result updates c and leads to an increment of the batch size; and second, the oldest fusion result is deleted from c , thus C is decremented again. Eventually, C remains constant. Due to considering and dropping of different c -values the counts c_i may become negative. Multiplication of the current c -vector with the batch size C helps to overcome this.

V. EXPERIMENTAL RESULTS

The proposed algorithm for updating the a-priori probability distribution of an unknown, stochastic, non-stationary process is analyzed in three experiments with different process characteristics with regard to accuracy. For this purpose two weakly coupled sensors with likelihoods of different accuracies are fused to obtain the optimal estimate \hat{x} .

In paragraph V-A the experimental setup is given. Afterwards, in V-B to V-D the experiments are presented. Finally, in subsection V-E the adaptive fusion results with regard to the non-stationary signals are evaluated.

A. Setup

- Experiments: We are interested in evaluating the adaption of the non-stationary a-priori probability functions for different signal behaviors: Alike in [27] different non-stationary signals are chosen, i.e. (i) abruptly changing (saltus function with infinite ascent/decent), (ii) linearly increasing/decreasing (sawtooth function) and (iii) a mixture of saltus and sawtooth functions.
- Unknown process X : Non-stationary, stochastic processes are generated that are represented by four realizations $X = \{x_1, x_2, x_3, x_4\} = \{1, 2, 3, 4\}$.
- A-priori probability function: The processes' realizations x are chosen at random by the Rejection method [29] following the given probability quantities of the generated process for each experiment.
- Measurement processes $Z_{(1)}$ and $Z_{(2)}$ observing X : The measurement processes shall be characterized by the same discrete realization spaces as X , i.e. $Z_{(1)} = Z_{(2)} = X$. Their sensor likelihoods are given by their conditional probability densities according to eq. (23) with $\pi \in [0.97, 0.91, 0.82, 0.73, 0.64, 0.55]$. In case of $\pi = 0.97$ the sensor can be called "highly accurate" while $\pi = 0.55$ would mean "intolerably inaccurate".
- Parameter values for learning process: We obtained (almost) optimal values for T_{KL} and C by grid optimization after running the fusion process 100 times.

- Estimator: We applied the MAP estimator.
- Evaluation: We computed the mean accumulated absolute mean error between x and \hat{x} , i.e. $\bar{E}(x, \hat{x}) = \frac{1}{n} \sum_{i=1}^n |x - \hat{x}|$ and its standard deviation $D(x, \hat{x})$. For quantification of the change between stationary and adaptive data fusion we computed the metrics $\Delta \bar{E}$ and ΔD .

$$P(z_{(i)}|x, \theta_{(i)}) = \begin{matrix} & \begin{matrix} z_{(i),1} & z_{(i),2} & z_{(i),3} & z_{(i),4} \end{matrix} \\ \begin{matrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{matrix} & \begin{bmatrix} \pi & \frac{1-\pi}{3} & \frac{1-\pi}{3} & \frac{1-\pi}{3} \\ \frac{1-\pi}{3} & \pi & \frac{1-\pi}{3} & \frac{1-\pi}{3} \\ \frac{1-\pi}{3} & \frac{1-\pi}{3} & \pi & \frac{1-\pi}{3} \\ \frac{1-\pi}{3} & \frac{1-\pi}{3} & \frac{1-\pi}{3} & \pi \end{bmatrix} \end{matrix} \quad (23)$$

B. Experiment 1

Let the a-priori probabilities of two realizations x_2 and x_3 be characterized by the saltus function (square function in Python) and $x_1 = x_4 = 0.05$ be constant:

$$P(x|u_X) = \begin{cases} 0.05 \\ 0.35 \cdot \text{Square}(t) + 0.5 \\ -0.35 \cdot \text{Square}(t) + 0.4 \\ 0.05 \end{cases} \quad (24)$$

It is obvious that the probabilities $P(x_2)$ and $P(x_3)$ abruptly change over time from higher probabilities to lower ones and vice-versa. The true a-priori probabilities $P(x_1), \dots, P(x_4)$ are shown as dashed lines in fig. 2(a) as well as the estimated probabilities of variables $P(\hat{x}_i)$ of X after applying adaptive data fusion with $T_{KL} = 0.01$ and $C = 30$ and the sensor likelihood in eq. (23) with $\pi = 0.91$. Note that $P(x_1)$ is not visible, since it is overplotted by $P(x_4)$, which has the same value. An impression of how the adaption to the true a-priori probabilities, particularly of the saltus functions by $P(\hat{x}_2)$ and $P(\hat{x}_3)$ by the proposed algorithm works, can be clearly identified. We can state that it takes some amount of time to adapt to true a-priori probabilities, because the algorithm starts with an equally distributed process $X|U_X$ (see for instance the increase of $P(\hat{x}_2)$ at the beginning). Also, the true probabilities may be under- or overestimated, particularly in case of the rare realizations x_1 and x_4 . The under- or overestimation can be explained by the small batch size and randomly chosen values for all x_i . Increasing C does not solve this problem, because of increasing process inertia. Decreasing the batch size further can lead to stronger fluctuations.

C. Experiment 2

Let the two realizations x_2 and x_3 be characterized by the a sawtooth function and $x_1 = x_4 = 0.05$ be constant:

$$P(x|u_X) = \begin{cases} 0.05 \\ 0.35 \cdot \text{Sawtooth}(t) + 0.5 \\ -0.35 \cdot \text{Sawtooth}(t) + 0.4 \\ 0.05 \end{cases} \quad (25)$$

In contrast to experiment 1 the probabilities of the true realizations $P(x_2)$ and $P(x_3)$ do not change abruptly, but take some time to increase/decrease linearly. The a-priori

probabilities $P(x_1), \dots, P(x_4)$ are shown as dashed lines in fig. 2(b). Additionally, the obtained a-priori probabilities of the non-stationary process in experiment 2 are depicted after applying adaptive data fusion with $T_{KL} = 0.01$ and $C = 25$ and the sensor likelihood in eq. (23) with $\pi = 0.91$. The labeling is the same as in experiment 1. In case of the “sawtooth” character of $P(\hat{x}_2)$ and $P(\hat{x}_3)$ it seems the adaption of the estimated probabilities is quicker than in case of the saltus function in experiment 1. However, there are again frequent under- and overestimations of the true probabilities.

D. Experiment 3

Let X be characterized by a non-stationary process, which is a mixture of the functions used in the experiments 1 and 2 with the following details:

$$P(x|u_X) = \begin{cases} 0.05 \\ 0.1 \cdot \text{Square}(t) + 0.4 \\ -0.1 \cdot \text{Sawtooth}(t) + 0.3 \\ 1.0 - P(x_1) - P(x_2) - P(x_3) \end{cases} \quad (26)$$

Clearly, the probabilities of the realizations $x_2 \dots x_3$ do not change dramatically and they have similar values. In fig. 2(c) the obtained variables of the non-stationary process in experiment 3 are depicted after applying adaptive data fusion with $T_{KL} = 0.025$ and $C = 75$ and the sensor likelihood in eq. (23) with $\pi = 0.91$. Apparently, due to their similarity in the probability domain, accurately estimating the realizations and distinguishing all realizations is a real challenge. Therefore, overestimations, particularly for $P(x_1)$ and $P(x_4)$, and underestimations, particularly for $P(x_2)$ and $P(x_3)$ occur frequently.

As shown in the figs. 2(a) and (b) of the experiments 1 and 2 the adaption to the underlying process works qualitatively fine in case accurate sensors are used, e.g. $\pi = 0.91$ in eq. (23). Apparently, in case experiment 3 (see fig. 2(c)) more accurate sensors are needed to improve the adaption process. The quantitative behavior will be analyzed in the following paragraph.

E. Evaluation of adaptive vs. stationary data fusion

The proposed algorithm will be analyzed for different sensor combinations and their corresponding likelihoods with different π -values in eq. (23) with regard to its behavior to stationary data fusion. For this purpose we assume that the stationary a-priori probabilities of all signals in eqs. (24), (25) and (26) are known to be capable of comparing the fusion results of adaptive data fusion with stationary data fusion.

In table I the results for all experiments are shown. In the first two columns the applied sensors are characterized by their π -values according to eq. (23). In the remaining columns characterized by $\Delta \bar{E}$ and ΔD the change in comparison to stationary data fusion is shown (in case of positive values adaptive data fusion is superior to stationary data fusion and vice versa). The following statements can be made:

- In case of abruptly over time changing a-priori probabilities (saltus functions) in experiment 1 the obtained

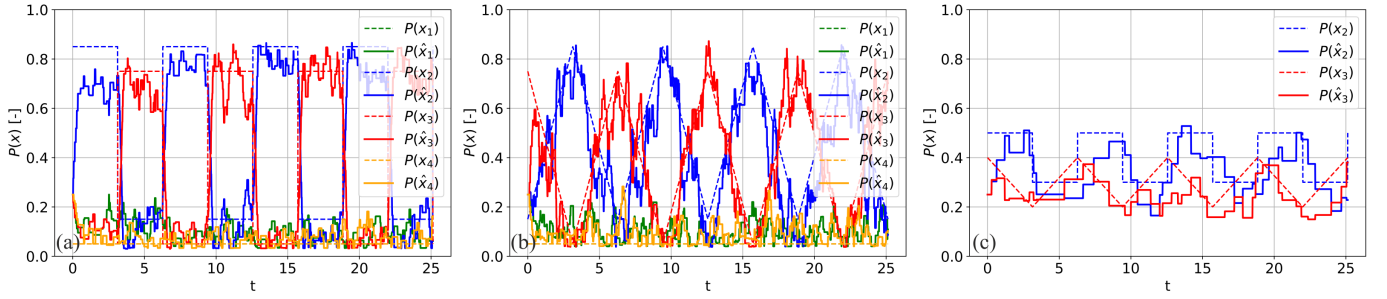


Fig. 2. Adaptation of the estimated, but unknown a-priori probability densities $P(\hat{x}_i)$ to the true ones $P(x_i)$ by data fusion: (a) saltus function experiment 1, (b) sawtooth/ramp function experiment 2, (c) “mixed” function experiment 3 (for reasons of clarity only the probabilities for x_2 and x_3 are shown).

accuracy can be improved by approximately 20%. Even if the sensors are not that accurate (e.g. $\pi \leq 0.72$) the adaptive version is in general better than the stationary. This finding is also true for experiment 2 in case of linearly over time changing probabilities if homogeneous sensors are applied. But these improvements reach a maximum of approximately 2%. In case of experiment 3 stationary data fusion is superior to adaptive fusion independently of the sensors applied.

- Although it is expected that any additional sensor improves the accuracy and reliability of the resulting data (see [15] in case of unbiased sensors), it is not the case for adaptive fusion. The combination of an accurate sensor (e.g. $\pi \in [0.97, 0.91]$) with a less accurate sensor (e.g. $\pi \leq 0.64$) results in a decrease of accuracy.
- We can, to some extent, state that the win in accuracy is a loss in deviation of the results, which is a cause of the small batch sizes in comparison to stationary data fusion. The more accurate both sensors are the smaller the deviation gets.
- The higher the batch size C and the higher T_{KL} the smoother the adaptive a-priori probabilities get. Higher T_{KL} values avoid an immediate adaption to the process change, which may be advantageous in case of noisy data, but it can be also a disadvantage due to an avoided adaption to the true process values.
- The more challenging an underlying process is (e.g. experiment 3), the bigger C seems to be, which we can also see in the overall batch sizes in experiments 2 and 3.
- It seems to be reasonable not to initialize the a-priori process variables equally distributed, because it takes too long to adapt to the real process. Apparently, (almost) any other a-priori probabilities seem to be more suitable.

VI. CONCLUSIONS AND FUTURE WORK

In this paper an algorithm is introduced that allows to determine time-dependent non-stationary a-priori probability density functions of an unknown, stochastic and non-stationary process. For this purpose, Bayesian Networks are applied to model cause-effect relationships between the unknown process and the measurement processes of the applied sensors as well as their weakly coupling in the sense of data fusion. Fur-

ther, Dirichlet distributions are applied for adaptive parameter learning of the unknown a-priori probabilities. The proposed algorithm is tested on three synthetic non-stationary processes of different characteristics as well as on combinations of sensors of different accuracies.

The results show that in case the a-priori probability distributions change abruptly (saltus/jump function) the accuracy of data fusion can be improved by up to approximately 20%, whereas in case of linearly changing process variables (ramp/sawtooth function) the improvements are marginal at approximately 2%. In contrast, in case the process variables are characterized by a mixture of ramp and saltus functions non-stationary data fusion is inferior to stationary data fusion. A possible interpretation is that the more the a-priori probabilities of the unknown process differ from each other (the more non-stationary a process really is) the better for adaptive data fusion, whereas the more similar (the more “equally distributed”) the a-priori probabilities of the unknown process variables are, the worse for adaptive data fusion.

Further, it is quite interesting that specifically improvements are made in case homogeneous sensors (or sensors that provide similar accuracies) are used. Although, usually each additional sensor (in case of an unbiased sensor it almost does not matter how good it is [15]) contributes by improving the fused data in the stationary case; however, in terms of adaptive, non-stationary data fusion it does matter, which types and accuracies of sensors are combined.

Clearly, these findings are only valid for the analyzed synthetic processes and the applied sensors; general statements cannot be made. Therefore, our future work is concerned with extending the analysis to more sophisticated process signal types with a greater variety of sensors. Although in previous investigations it was shown that measurement data of biased sensors can be de-biased it seems reasonable to apply the proposed methods on accurate, but strongly biased sensors as well. Further, there seems to be an “accuracy border” between abruptly changing signals (infinite ascents/descents) and linearly changing signals (finite ascents/descents), which we want to quantify. Also, the batch size C for collection knowledge about the preceding fusion results remained constant for each single process. However, non-stationarity can change over time rather quickly in different ways. Therefore

TABLE I
ADAPTIVE VS. STATIONARY DATA FUSION (NOTE THAT POSITIVE ΔE AND ΔD VALUES SHOW AN IMPROVEMENT AND VICE VERSA).

		Experiment 1						Experiment 2						Experiment 3					
π	S1	C_{opt}	$T_{KL,opt}$	E	$\Delta E[\%]$	D	$\Delta D[\%]$	C_{opt}	$T_{KL,opt}$	E	$\Delta E[\%]$	D	$\Delta D[\%]$	C_{opt}	$T_{KL,opt}$	E	$\Delta E[\%]$	D	$\Delta D[\%]$
0.97	0.97	30	0.01	32.7	22.2	7.0	12.6	30	0.01	32.7	1.8	7.1	3.2	50	0.025	57.5	-6.0	10.5	4.7
0.97	0.91	20	0.05	46.4	11.0	9.0	0.8	30	0.01	58.3	-13.5	10.6	-17.2	150	0.05	88.4	-14.2	12.7	-18.9
0.97	0.82	25	0.01	72.0	3.6	10.1	-2.6	20	0.001	78.3	-8.3	10.0	-0.7	10	0.0025	91.9	-2.3	11.2	5.8
0.97	0.73	25	0.01	82.7	1.1	12.8	-0.6	100	0.5	86.5	-2.1	11.6	-12.6	200	0.05	92.0	0.0	12.9	0.0
0.97	0.64	25	0.001	83.7	-2.3	11.3	-3.1	20	0.01	84.0	0.0	11.6	0.3	50	0.05	91.2	-0.2	12.1	0.0
0.97	0.55	30	0.01	83.7	-0.3	11.0	1.3	30	0.01	83.6	-0.3	12.7	-0.6	20	0.5	93.0	0.0	13.4	0.0
0.91	0.91	15	0.001	88.5	20.8	12.0	9.5	35	0.005	110.9	2.1	12.9	4.6	75	0.0025	178.6	-3.9	17.0	-11.0
0.91	0.82	30	0.005	133.1	9.0	14.5	-2.9	25	0.01	157.4	-7.7	18.4	-23.0	35	0.2	235.5	-7.2	19.8	0.9
0.91	0.73	30	0.001	168.7	3.9	17.4	-18.7	30	0.0025	194.0	-12.6	19.2	-13.4	75	0.2	263.3	-8.7	23.4	-17.2
0.91	0.64	35	0.01	209.5	-1.8	17.5	-7.3	30	0.01	228.1	-11.8	17.7	-4.3	50	0.05	276.5	-5.0	20.8	3.4
0.91	0.55	25	0.0025	235.7	0.0	19.2	-19.9	25	0.01	246.4	-4.9	20.1	1.4	5	0.0025	275.5	-0.9	22.1	-4.4
0.82	0.82	30	0.0025	249.0	12.1	16.6	12.8	35	0.001	240.9	1.1	18.1	-7.4	35	0.1	376.3	-5.3	24.5	0.9
0.82	0.73	20	0.025	295.6	7.2	20.3	3.0	35	0.001	315.2	-5.6	21.2	-3.0	35	0.1	453.1	-4.0	29.4	-35.5
0.82	0.64	25	0.0025	342.0	4.8	24.7	-11.1	30	0.0025	368.1	-9.6	25.0	-12.0	35	0.2	493.3	-7.0	33.9	-25.3
0.82	0.55	25	0.005	382.8	0.0	24.3	-19.9	100	0.2	427.0	-12.3	27.8	-26.4	50	0.1	535.3	-7.6	30.7	-14.4
0.73	0.73	30	0.0025	353.4	6.2	25.0	-29.2	30	0.005	410.1	-7.2	26.2	-25.3	35	0.1	587.5	-5.6	35.4	-25.4
0.73	0.64	25	0.01	411.2	5.2	23.5	-13.8	35	0.0025	489.9	-10.6	28.5	-39.2	35	0.1	674.3	-2.9	37.7	-12.6
0.73	0.55	25	0.01	478.2	3.6	26.5	-20.4	25	0.2	550.9	-12.4	48.7	-131.0	30	0.1	738.5	-6.0	39.5	-33.7
0.64	0.64	30	0.001	482.2	9.5	27.6	-20.3	30	0.0025	566.6	-5.6	31.1	-32.8	30	0.1	821.9	-6.8	39.7	-33.4
0.64	0.55	20	0.01	547.0	9.9	31.1	-20.4	25	0.0025	650.8	-7.2	32.4	-28.9	35	0.05	923.2	-4.1	53.7	-59.0
0.55	0.55	25	0.001	594.4	16.9	37.2	-51.1	25	0.001	721.5	-0.9	38.1	-46.1	150	0.0025	1082.0	-8.1	55.9	-71.2

it seems reasonable to permanently adapt the batch size to the needs of the underlying process and the applied sensors.

REFERENCES

- [1] K. Fang, *Track-RNN: Joint Detection and Tracking Using Recurrent Neural Networks*, 29th Conference on Neural Information Processing Systems, Barcelona, Spain, 2016.
- [2] A. Krizhevsky, I. Sutskever, and G. Hinton, *Imagenet classification with deep convolutional neural networks*. In *Advances in Neural Information Processing Systems* 25, 2012.
- [3] M. Junghans, and H.J. Jentschel, *Qualification of traffic data by bayesian network data fusion.*, 10th International Conference on Information Fusion, Québec, Canada, July 2007.
- [4] M. Junghans, and H.J. Jentschel, *Correction of Selection Bias in Traffic Data by Bayesian Network Data Fusion*, *Journal of Advances in Information Fusion*, Vol. 3, No. 1, pp. 50-62, June 2008.
- [5] D. Heckerman, *A tutorial on learning with Bayesian networks*, Technical Report MSRTR-95-06, <http://www.cis.upenn.edu/~mkearns/papers/barbadosheckerman.pdf>, Microsoft Research, 1996.
- [6] R. Neapolitan, *Learning Bayesian Networks*. Pearson Education, Inc., Prentice Hall, 2004.
- [7] D. Spiegelhalter, and S. Lauritzen, *Sequential Updating of Conditional Probabilities on Directed Graphical Structures*, *Networks*, Vol. 20, pp. 579-605, 1990.
- [8] R.G. Cowell, A.P. Dawid, S.L. Lauritzen, D.J. Spiegelhalter, *Probabilistic Networks and Expert Systems*. Springer Verlag New York, Inc., 1999.
- [9] H. Herold, *Sensortechnik. Sensorwirkprinzipien und Sensorsysteme*, Hüthig Buch Verlag Heidelberg, 1993.
- [10] N. El Faouzi, and L.A. Klein, *Data Fusion for ITS: Techniques and Research Needs*, *Transportation Research Procedia* Vol. 15, 2016.
- [11] D.L. Hall, *Mathematical Techniques in Multisensor Data Fusion*, Artech House, Norwood, Massachusetts, USA, 1992.
- [12] L.A. Klein, *Sensor and Data Fusion: A Tool for Information Assessment and Decision Making*, SPIE Press, Bellingham (WA), 2004.
- [13] J. Llinas, E.L. Waltz, *Multisensor Data Fusion*. Artech House, Inc., Norwood, Massachusetts, 1990.
- [14] D.P. Mandic, D. Obradovic, A. Kuh, T. Adali, U. Trutschel, M. Golz, P. de Wilde, J. Barria, A. Constantinides, J. Chambers, *Data Fusion for Modern Engineering: An Overview*, In *Lecture Notes in Computer Science*, Vol. 3697, pp. 715-721, Springer Verlag Berlin Heidelberg, 2005.
- [15] T. Neumann, R. Ebendt, and G. Kuhns, *From Finance to ITS: traffic data fusion based on Markowitz portfolio theory*, *Journal of Advanced Transportation* Vol. 50, pp. 145-164, 2015.
- [16] L. Wald, *Some Terms of Reference in Data Fusion*, In *IEEE Transactions on Geoscience and Remote Sensing*. Vol. 37, No. 3, 1999.
- [17] A. Berler, S.E. Shimony, *Bayes Networks for Sonar Sensor Fusion*, in D. Geiger, and P. Shenoy (eds), *Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence*, 1997.
- [18] C. Cappelle, M.E. El Najjar, and D. Pomorski, *Multi-sensors data fusion using dynamic bayesian network for robotised vehicle geo-localisation*, 11th International Conference on Information Fusion, 30 June-3 July 2008, Cologne, Germany, 2008.
- [19] Y.M. Chen, C. Hsueh, and C. Wang, *Data Mining of Bayesian Networks to Select Fusion Nodes from Wireless Sensor Networks*, *IJCSI International Journal of Computer Science Issues*, Volume 13, Issue 4, July 2016.
- [20] A. De Paola, S. Gaglio, G.L. Re, and M. Ortolani, *Multi-sensor fusion through adaptive bayesian networks*, In: R. Pirrone, F. Sorbello (eds) *AI*IA 2011: Artificial Intelligence Around Man and Beyond*. Lecture Notes in Computer Science, vol. 6934. Springer, Berlin, Heidelberg, 2011, DOI: https://doi.org/10.1007/978-3-642-23954-0_33.
- [21] M. Junghans, and A. Leich, *Traffic State Estimation with Bayesian Networks at Extremely Low V2X Penetration Rates*, 19th International Conference on Information Fusion, Heidelberg, Germany, July 2016.
- [22] C. Smaili, M.E. El Najjar, and F. Charpillat, *Multi-Sensor Fusion Method using Dynamic Bayesian Network for Precise Vehicle Localization and Road Matching*, 19th IEEE International Conference on Tools with Artificial Intelligence, vol. 1, 2007, pp. 146-151.
- [23] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Second Printing, Morgan Kaufmann Publishers, Inc., San Francisco (CA), USA, 1991.
- [24] J.J. Clark, and A.L. Yuille, *Data Fusion for Sensory Information Processing Systems*, Kluwer Academic Publishers Boston/London, 1990.
- [25] I. Cohen, A. Bronstein and F.G. Cozman, *Adaptive Online Learning of Bayesian Network Parameters*, <https://pdfs.semanticscholar.org/5286/23a13893e01414ae8180ad6de464cccd73af.pdf>, 2001.
- [26] J. Liu, Q. Liao, *Online Learning of Bayesian Network Parameters*, 2008 Fourth International Conference on Natural Computation, Jinan, Shandong, China, 18-20 October 2008.
- [27] M. Junghans, and H.J. Jentschel, *Adaptive Learning of Bayesian Networks for the Qualification of Traffic Data by Contaminated Dirichlet Density Functions*, *Proceedings of the 2008 1st International Conference on Information Technology*, Gdansk, Poland, May 2008.
- [28] M. Junghans, *Ein Beitrag zur Qualifizierung von Verkehrsdaten mit Bayesschen Netzen*, Ph.D. thesis, Dresden University of Technology, Dresden, Germany, 2009.
- [29] J. von Neumann, *Various techniques used in connection with random digits*, *Collected Works*, vol. 5, pp. 768-770. Pergamon Press, 1963.